

Challenges to Standardization: A Case Study Using Coastal and Deep-Ocean Water Level Data

Aaron Sweeney¹, Kelly Stroker¹, George Mungov¹, Susan McLean²

¹Cooperative Institute for Research in Environmental Sciences, Boulder, CO, ²National Centers for Environmental Information, National Oceanic and Atmospheric Administration, Boulder, CO, United States

Introduction

Sea levels recorded at coastal stations and inferred from deep-ocean pressure observations at the seafloor are submitted for archive in multiple data and metadata formats. These formats include two forms of schema-less XML and a custom binary format accompanied by metadata in a spreadsheet. The authors report on efforts to use existing standards to make this data more discoverable and more useful beyond their initial use in detecting tsunamis. An initial review of data formats for sea level data around the globe revealed heterogeneity in presentation and content. In the absence of a widely-used domain-specific format, we adopted the general model for structuring data and metadata expressed by the **Network Common Data Form (netCDF)**. netCDF has been endorsed by the **Open Geospatial Consortium** and has the advantages of small size when compared to equivalent plain text representation and provides a standard way of embedding metadata in the same file. We followed the orthogonal time-series profile of the **Climate and Forecast discrete sampling geometries** as the convention for structuring the data and describing metadata relevant for use. We adhered to the **Attribute Convention for Data Discovery** for capturing metadata to support user search. Beyond making it possible to structure data and metadata in a standard way, netCDF is supported by multiple software tools in providing programmatic cataloging, access, subsetting, and transformation to other formats. We will describe our successes and failures in adhering to existing standards and provide requirements for either augmenting existing conventions or developing new ones. Some of these enhancements are specific to sea level data, while others are applicable to time-series data in general.

Data collection not designed around re-use...

Coastal Sea Level Time-series Data, as sent to NCEI (Provider 1)

- No XML schema for data (non-standard content)
- Reference datum, platform description, georeference, sensor information, and units supplied in separate file

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<metadata>
  <idinfo>
    <datasetid> gov.noaa.ngdc.1min_tsunamiWL: 1611400 </datasetid>
    <citation>
      <citeinfo>
        <origin>DOC/NOAA/NOS/CO-OPS &gt; Center for Operational Oceanographic Products and Services,
        National Ocean Service, NOAA, U.S. Department of Commerce </origin>
        <pubdate> 20150505 </pubdate>
        <title>CO-OPS 1-minute Raw Tsunami Water level Data for StationId 1611400 start date 20150504 to
        end date 20150504 </title>
        <geoform>map </geoform>
        <pubinfo>
          <pubplace>Silver Spring, MD </pubplace>
          <publish>NOAA&apos;s Ocean Service, Center for Operational Oceanographic Products and Services
          (CO-OPS) </publish>
        </pubinfo>
        <onlink> http://tidesandcurrents.noaa.gov/<onlink>
        <onlink>http://opendap.co-ops.nos.noaa.gov/stations/index.jsp</onlink>
        ...
      </citeinfo>
    </idinfo>
  </metadata>
```

```
<?xml version="1.0" encoding="utf-8" ?>
<tsunamiData>
  <station_id>1611400</station_id>
  <station_name>NAWILIWILI, NAWILIWILI HARBOR, KAUAI</station_name>
  <state>HI</state>
  <sensor_id>U1</sensor_id>
  <dcp>1</dcp>
  <data_source>USDOC/NOAA/NOS/COOPS(Center for Operational Oceanographic Products and
  Services)</data_source>
  <COOPS_disclaimer>These raw data have not been subjected to the National Ocean Service's quality control
  or quality assurance procedures and do not meet the criteria and standards of official National Ocean Service
  data. They are released for limited public use as preliminary data to be used only with appropriate
  caution.</COOPS_disclaimer>
  <tsunamiWaterLevelData>
    <dateTime>2015-05-04T00:00:00Z</dateTime>
    <Val_1_MIN>1.143</Val_1_MIN>
    <dateTime>2015-05-04T00:01:00Z</dateTime>
    <Val_1_MIN>1.142</Val_1_MIN>
    <dateTime>2015-05-04T00:02:00Z</dateTime>
    <Val_1_MIN>1.149</Val_1_MIN>
    ...
  </tsunamiWaterLevelData>
</tsunamiData>
```

Coastal Sea Level Time-series Data, as sent to NCEI (Provider 2)

- No XML schema (non-standard content)
- No reference datum, platform, sensor, and georeference details supplied in file
- Files created every 15 minutes
- Data from all stations in single file

```
<TideData>
<Transmission>
  <OperatorID>1</OperatorID>
  <UUID>00646cc3-3c96-4cd8-a511-c3c3058dc1e4</UUID>
  <DateTime>2015-12-08T18:13:02UTC</DateTime>
</Transmission>
<TideRecord>
  <StationID operator="1">9411166</StationID>
  <WaterLevel unit="centimeters">151.18</WaterLevel>
  <DateTime>2015-12-08T17:51:15UTC</DateTime>
</TideRecord>
<TideRecord>
  <StationID operator="1">9411166</StationID>
  <WaterLevel unit="centimeters">152.1</WaterLevel>
  <DateTime>2015-12-08T17:51:30UTC</DateTime>
</TideRecord>
<TideRecord>
  <StationID operator="1">9411166</StationID>
  <WaterLevel unit="centimeters">151.79</WaterLevel>
  <DateTime>2015-12-08T17:51:45UTC</DateTime>
</TideRecord>
...
```

Ocean Bottom Pressure Time-series Data, as sent to NCEI (Provider 3)

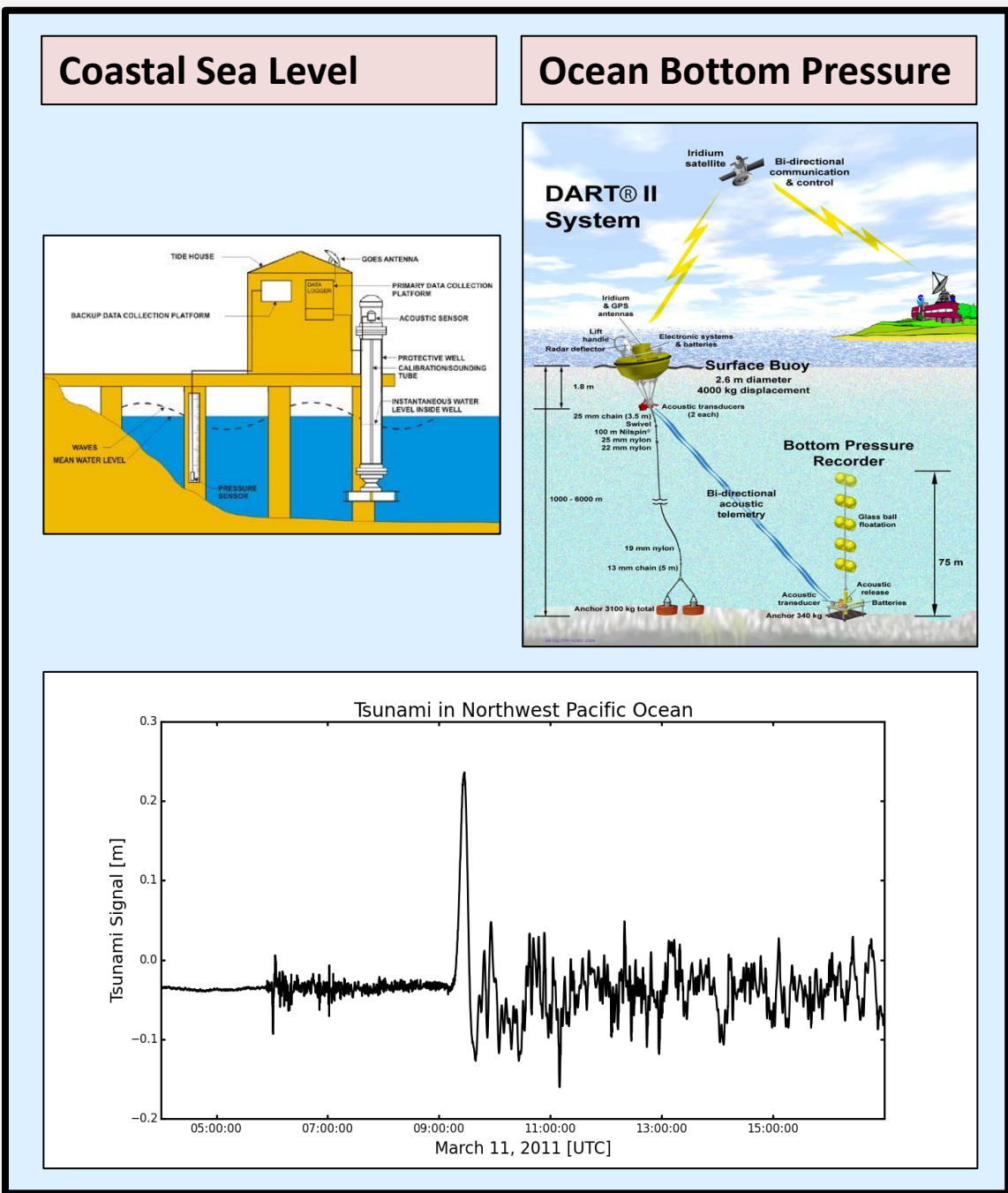
- Non-standard binary format
- Content is engineering data (not physical data)
- Calibration information provided separately
- Platform, sensor, and georeference details provided separately (Excel spreadsheet)

Records consist of 2048 byte blocks. Each block consists of 30 header bytes, followed by 1920 data bytes, and 98 unused bytes.

Bytes	Meaning
1-2	CAFÉ start header bytes
3-4	checksum (computed on data bytes only)
5-8	date
9-11	time
12	battery voltage
13	BPR serial number
14-17	pressure oscillator serial number (ParoScientific)
18	reference oscillator serial number
19-26	reference oscillator period
27	pressure frequency divider
28	temperature frequency divider
29	temperature heterodyned frequency divider
30	sample rate
31-1950	data (one hour of 15-second data)
1951-2048	unused (but required by flash memory)

Each 15-second data reading consists of eight bytes as follows:

Bytes	Meaning
1-2	pressure counts
3-4	pressure vernier counts
5-6	temperature counts
7-8	temperature vernier counts



netCDF promotes data discovery and re-use...

- Coastal and Deep-Ocean Water Level Data match the structure of netCDF orthogonal time-series
- netCDF metadata can be easily transformed to ISO-compliant XML, which drives discovery through data.noaa.gov and data.gov
- netCDF is easily converted to other formats (CSV, XML, WaterML2)
- Platform, sensor, georeference, and reference datum details, essential for data re-use, travel with data

Coastal Sea Level Time-series – Information Supporting Data Re-Use

Physical Variables:

waterlevel_unassessed:long_name = "Water Level Relative to Station Datum" ;
waterlevel_unassessed:standard_name = "water_surface_height_above_reference_datum" ;
waterlevel_unassessed:coverage_content_type = "physicalMeasurement" ;
waterlevel_unassessed:units = "m" ;
waterlevel_unassessed:ancillary_variables = "waterlevel_datum_altitude" ;
waterlevel_unassessed:platform = "platform_info" ;
waterlevel_unassessed:instrument = "sensor_info" ;

waterlevel_datum_altitude:long_name = "Altitude is the (geometric) height above the geoid, which is the reference geopotential surface. The geoid is similar to mean sea level. Water_surface_reference_datum_altitude means the altitude of the arbitrary datum referred to by a quantity with standard name water_surface_height_above_reference_datum." ;
waterlevel_datum_altitude:standard_name = "water_surface_reference_datum_altitude" ;
waterlevel_datum_altitude:coverage_content_type = "physicalMeasurement" ;
waterlevel_datum_altitude:units = "m" ;

Reference Datum:

:geospatial_vertical_reference = "Station Datum" ;

Georeference:

lat = 21.95444 ;
lat:units = "degree_north" ;
lon = -159.3561 ;
lon:units = "degree_east" ;
crs:epsg_code = "EPSG:4326" ; (i.e. WGS-84)

Platform Details

platform_info:long_name = "Tide Station 1611400, NAWILIWILI, NAWILIWILI HARBOR, KAUAI" ;
platform_info:comment = "1954-11-24: NOAA NOS CO-OPS establishment of Station 1611400 in support of the NWLON, 2005-06-14: Station 1611400 Sensor and DCP updated to receive 1 minute water level data" ;
platform_info:ncei_code = "TIDE STATIONS" ;

Sensor Details

sensor_info:long_name = "Tide Gauge" ;
sensor_info:precision = "0.001" ;
sensor_info:accuracy = "0.009" ;
sensor_info:units = "m" ;
sensor_info:type = "Acoustic Sensor" ;
sensor_info:make = "Aquatrak" ;
sensor_info:model = "NG-3000 Water Level Sensor" ;
sensor_info:serial_number = "" ;
sensor_info:coops_sensor_id = "U1" ;
sensor_info:coops_dcp_id = "1" ;
sensor_info:sensor_history = "2015-01-01: Type Acoustic Sensor Make Aquatrak Model NG-3000 Water Level Sensor Serial Number 0941-3218 Sensor ID U1 DCP 1 Accuracy 0.009 m, 2015-02-22: Type Acoustic Sensor Make Aquatrak Model NG-3000 Water Level Sensor Serial Number 2264-AQ0174 Sensor ID U1 DCP 1 Accuracy 0.009 m" ;

Citation Information:

:metadata_link = "http://doi.org/10.7289/V59884XF" ;

Deficiencies

- The value for "water_surface_reference_datum_altitude," i.e. the height of the reference datum above the geoid, is not currently reported with the sea level data and must be obtained from geodetic surveys.
- Translations between local tidal datums are not captured.
- The orthogonal time-series model does not provide a way to track changes in station location over time.

Ocean Bottom Pressure Time-series – Information Supporting Data Re-Use

Physical Variables:

seafloor_pressure_abs_raw:long_name = "Observed Absolute Pressure at Sea Floor" ;
seafloor_pressure_abs_raw:standard_name = "sea_water_pressure_at_sea_floor" ;
seafloor_pressure_abs_raw:coverage_content_type = "physicalMeasurement" ;
seafloor_pressure_abs_raw:units = "dbar" ;
seafloor_pressure_abs_raw:platform = "buoy_info, bottom_pressure_recorder_info" ;
seafloor_pressure_abs_raw:instrument = "pressure_sensor_info" ;
seafloor_pressure_abs_raw:comment = "1. A BPR was deployed at 21418 on 2013-07-13, but the system spiked. It was recovered and a new BPR deployed. 2. Real-time system failed 2014-05-03. 3. Lab RFO record could not be located." ;

Georeference:

lat = 38.69000 ;
lat:units = "degree_north" ;
lon = 148.80306 ;
lon:units = "degree_east" ;
crs:epsg_code = "EPSG:4326" ; (i.e. WGS-84)

Platform Details

bottom_pressure_recorder_info:long_name = "DART Bottom Pressure Recorder (BPR) Deployment 21418_20130715to20140908" ;
bottom_pressure_recorder_info:comment = "The bottom pressure recorder includes a timing circuit based on a reference frequency oscillator (rfo). The period of the rfo is typically measured before and after deployment to monitor drift." ;
bottom_pressure_recorder_info:type = "DART II" ;
bottom_pressure_recorder_info:wmo_code = 21418 ;
bottom_pressure_recorder_info:bpr_board_serial_number = 9004994 ;
bottom_pressure_recorder_info:software_version = "2.78" ;
bottom_pressure_recorder_info:trigger_threshold = 30 ;
bottom_pressure_recorder_info:trigger_threshold_uom = "mm" ;
bottom_pressure_recorder_info:clock_check_gps_pre_deployment = "2013-07-14T23:39:00Z" ;
bottom_pressure_recorder_info:clock_check_bpr_pre_deployment = "2013-07-14T23:39:00Z" ;
bottom_pressure_recorder_info:clock_check_gps_post_recovery = "2014-09-08T20:56:00Z" ;
bottom_pressure_recorder_info:clock_check_bpr_post_recovery = "2014-09-08T20:56:28Z" ;
bottom_pressure_recorder_info:rfo_period = 0.476836949 ;
bottom_pressure_recorder_info:rfo_period_uom = "microsecond" ;
bottom_pressure_recorder_info:rfo_period_at_0_degC = 0.476836838 ;
bottom_pressure_recorder_info:rfo_period_at_0_degC_date = "2013-06-05T00:00:00Z" ;
bottom_pressure_recorder_info:rfo_period_pre_deployment = 0.476836852 ;
bottom_pressure_recorder_info:rfo_period_pre_deployment_date = "2013-07-14T00:00:00Z" ;
bottom_pressure_recorder_info:rfo_period_post_recovery = 0.476836838 ;
bottom_pressure_recorder_info:rfo_period_post_recovery_date = "" ;

Sensor Details:

pressure_sensor_info:long_name = "Pressure Sensor" ;
pressure_sensor_info:precision = 0.006 ;
pressure_sensor_info:accuracy = 0.6 ;
pressure_sensor_info:units = "dbar" ;
pressure_sensor_info:make = "Paroscientific, Inc." ;
pressure_sensor_info:model = "410K Digiquartz Pressure Sensor" ;
pressure_sensor_info:serial_number = 95795 ;
pressure_sensor_info:factory_calibrated = "true" ;
pressure_sensor_info:valid_range = 0., 6894. ;
pressure_sensor_info:comment = "An absolute pressure transducer with a Digiquartz pressure sensor and precision thermometer. Operates over a temperature range of 0-125 degrees Celsius and a pressure range of 0-10000 psia with a typical accuracy of 0.01% and a resolution of 0.0001%." ;
pressure_sensor_info:sdn_instrument_urn = "SDN:L22::TOOL0403" ;
pressure_sensor_info:sdn_instrument_name = "Paroscientific 410K Pressure Transducer" ;

Citation Information:

:metadata_link = "http://doi.org/10.7289/V5F18WNS" ;

References

NCEI netCDF Templates v2.0: www.nodc.noaa.gov/data/formats/netcdf/v2.0 (includes CF and ACDD guidance)
NCEI Water Level Data (netCDF): www.ngdc.noaa.gov/thredds/enhancedCatalogWaterLevel.html

Acknowledgements

The authors thank the NCEI netCDF Team, especially Ajay Krishnan and Matt Biddle, for providing a critical review of the resultant netCDF files .

